

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. Patent Application of

NAGASAKA *et al.*

Application Number: To Be Assigned

Filed: Concurrently Herewith

**For: METHOD AND APPARATUS FOR CHARACTER
STRING SEARCH IN IMAGE**



**Honorable Assistant Commissioner
for Patents
Washington, D.C. 20231**

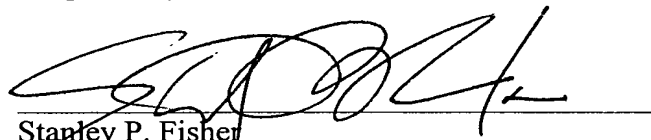
**REQUEST FOR PRIORITY
UNDER 35 U.S.C. § 119
AND THE INTERNATIONAL CONVENTION**

Sir:

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of March 22, 2001, the filing date of Japanese patent applications 2001-082012.

The certified copy of Japanese patent application 2001-082012 is being submitted herewith. Acknowledgment of receipt of the certified copies is respectfully requested in due course.

Respectfully submitted,


Stanley P. Fisher
Registration Number 24,344

REED SMITH HAZEL & THOMAS LLP
3110 Fairview Park Drive
Suite 1400
Falls Church, Virginia 22042
(703) 641-4200

JUAN CARLOS A. MARQUEZ
Registration No. 34,072

December 27, 2001

3/ 01-238 0:

5/11/02
6/1/02

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出 願 年 月 日
Date of Application:

2001年 3月22日

出 願 番 号
Application Number:

特願2001-082012

出 願 人
Applicant(s):

株式会社日立製作所

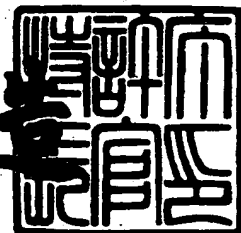


CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年10月19日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



【書類名】 特許願

【整理番号】 NT01P0171

【提出日】 平成13年 3月22日

【あて先】 特許庁長官 殿

【国際特許分類】 G06T 7/00

【発明者】

 【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所 中央研究所内

 【氏名】 長坂 晃朗

【発明者】

 【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所 中央研究所内

 【氏名】 宮武 孝文

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100068504

 【弁理士】

 【氏名又は名称】 小川 勝男

 【電話番号】 03-3661-0071

【選任した代理人】

 【識別番号】 100086656

 【弁理士】

 【氏名又は名称】 田中 恭助

 【電話番号】 03-3661-0071

【選任した代理人】

 【識別番号】 100094352

 【弁理士】

【氏名又は名称】 佐々木 孝

【電話番号】 03-3661-0071

【手数料の表示】

【予納台帳番号】 081423

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 映像中の文字検索方法及び装置

【特許請求の範囲】

【請求項 1】

映像を入力する工程と、入力した映像のフレーム画像から文字領域をその形状に基づいて検出する工程と、該文字領域の第一の画像特徴を抽出する工程と、文字入力手段によって入力された検索対象の文字列を画像として描画する工程と、描画した該文字列画像から第二の画像特徴を抽出する工程と、該第一の画像特徴と該第二の画像特徴とを照合して一致度を求める工程と、一致度が得られた文字列を含む該文字領域を出力する工程とを有することを特徴とする映像中の文字検索方法。

【請求項 2】

前記文字領域を出力する工程は、一致度の高い順に該文字領域を出力することを特徴とする請求項 1 に記載の映像中の文字検索方法。

【請求項 3】

前記文字領域をその形状に基づいて検出する工程は、予め指定する範囲の輝度の差の画素が予め指定する範囲の長さ分だけ連続して繋がっている等輝度画素列が局所的に集中して存在する領域を文字領域として判定することを特徴とする請求項 1 又は請求項 2 に記載の映像中の文字検索方法。

【請求項 4】

前記文字領域をその形状に基づいて検出する工程は、予め指定する範囲の輝度の差の画素が予め指定する範囲の長さ分だけ水平方向に連続して繋がっている水平等輝度画素列と、予め指定する範囲の輝度の差の画素が予め指定する範囲の長さ分だけ垂直方向に連続して繋がっている垂直等輝度画素列とが、同じ画像範囲内に同時に存在する領域を文字領域として判定することを特徴とする請求項 1 又は請求項 2 に記載の映像中の文字検索方法。

【請求項 5】

前記文字領域の第一の画像特徴を抽出する工程及び前記描画した該文字列画像から第二の画像特徴を抽出する工程は、抽出する該第一の画像特徴及び該第二の

画像特徴がいずれも、文字領域及び文字列を2値化することによって得られる垂直方向のエッジ数を水平方向に順に並べた1次元特徴列であり、前記一致度を求める工程は、該第一の画像特徴である該1次元特徴列と該第二の画像特徴である該1次元特徴列との伸縮照合によって一致度を求めることを特徴とする請求項1～請求項4のいずれかーに記載の映像中の文字検索方法。

【請求項6】

前記文字領域の第一の画像特徴を抽出する工程及び前記描画した該文字列画像から第二の画像特徴を抽出する工程は、抽出する該第一の画像特徴及び該第二の画像特徴がいずれも、文字領域及び文字列を2値化することによって得られる水平方向のエッジ数を垂直方向に順に並べた1次元特徴列であり、前記一致度を求める工程は、該第一の画像特徴である該1次元特徴列と該第二の画像特徴である該1次元特徴列との伸縮照合によって一致度を求めることを特徴とする請求項1～請求項4のいずれかーに記載の映像中の文字検索方法。

【請求項7】

映像を入力する手段と、入力したフレーム画像から文字領域を検出する手段と、該文字領域の第一の画像特徴を抽出する手段と、任意の文字コードを入力するための入力手段と、該文字コードに対応する文字を画像として描画する文字画像生成手段と、該文字画像から第二の画像特徴を抽出する手段と、該画像特徴と、先に入力フレーム画像から抽出した画像特徴とを照合し一致度を求める特徴照合手段と、一致度が得られた画像特徴に対応する文字領域又はそれを含むフレーム画像を検索結果として出力する出力手段とを有することを特徴とする映像中の文字検索装置。

【請求項8】

前記出力手段は、特徴照合手段によって得られた一致度の高い順に順位付けして検索結果を出力することを特徴とする請求項7に記載の映像中の文字検索装置。

【請求項9】

前記文字領域を検出する手段は、予め指定する範囲の輝度の差の画素が予め指定する範囲の長さ分だけ連続して繋がっている等輝度画素列が局所的に集中して

存在することを判定条件とすることを特徴とする請求項 7 又は請求項 8 に記載の映像中の文字検索装置。

【請求項 1 0】

前記文字領域の第一の画像特徴を抽出する手段及び前記文字画像から第二の画像特徴を抽出する手段は、抽出する該第一の画像特徴及び該第二の画像特徴がいずれも、文字領域及び文字列を 2 値化することによって得られる垂直方向のエッジ数を水平方向に順に並べた 1 次元特徴列であり、前記特徴照合手段は、該第一の画像特徴である該 1 次元特徴列と該第二の画像特徴である該 1 次元特徴列との伸縮照合によって特徴照合を行なうことを特徴とする請求項 7 ～請求項 9 のいずれか一に記載の映像中の文字検索装置。

【請求項 1 1】

コンピュータに、入力した映像のフレーム画像から文字領域をその形状に基づいて検出する手順と、該文字領域の第一の画像特徴を抽出する手順と、文字入力手段によって入力された検索対象の文字列を画像として描画する手順と、描画した該文字列画像から第二の画像特徴を抽出する手順と、該第一の画像特徴と該第二の画像特徴とを照合して一致度を求める手順と、一致度が得られた文字列を含む該文字領域を出力する手順とを実行させるための映像中の文字検索処理プログラム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、放送中の映像又はデータベース中の映像の検索技術に係り、特に映像中の文字を手掛かりとした検索が高速に行なえる映像の検索方法に関する。

【0 0 0 2】

【従来の技術】

映像において、字幕（画面中に嵌め込まれる文字、テロップとも云う）やフリップ（被写体として写される文字）などの文字情報は、それらが現れている場面の内容を端的に象徴する重要な情報の一つである。そのため、これら字幕が現れる箇所の検出や、字幕で使われている文字列テキストを対象にしたキーワード検

索などの研究が精力的に進められている。

【0003】

ここで、字幕の領域の検出に当たっては、(1) 文字を形作る線図形部分が高輝度であること、(2) その文字部分と境界との間の輝度差が高く、それによってエッジが形成されると共に、それらエッジが文字近傍に多数存在すること、(3) 文字領域は一定時間静止して存在すること、を前提とする手法が一般的である。また、キーワード検索のためには、上記で検出した文字領域を、文書処理で広く使われている文字認識によってASCII (American Standard Code for Information Interchange) やJIS (Japanese Industrial Standards) 等の文字コードに変換し、そのコードを基に照合を行なう方法がある。

【0004】

これとは別に、文字の形状に着目して検索を行なう手法が開示されている。例えば、1988年リアオ(RIAO)大会予稿集(Proceedings of Recherche d'Information Assistee par Ordinateur, Cambridge, MA)第248頁～第258頁(Y. Tanaka and H. Torii, "Transmedia machine and its keyword search over image texts")では、スキャナー等で読み取った印刷文書中から1文字1文字を切り出し、各々について形状特徴に基づくコードを決定して割り当てておき、一方、キーボード等から入力された文字列の各文字についても同様に同じコードを与えることで、それらコード間の照合を行なっている。

【0005】

【発明が解決しようとする課題】

上記文献の技術は、1文字毎にその形状に対してコードを対応づける点で、文字認識を簡略化したものと云える。この場合、一文字毎にどのようなコードを与えるかを前もって決める必要があり、同じ検索手法を他の言語に適用するためには、特徴量の割り当て方法を各言語に合わせて変える必要がある。

【0006】

また、1文字毎に特徴量を割り当てるために、1文字の切り出しが正しく行なわれることが前提となるが、無地の背景に文字をコントラスト高く印刷した文書等以外では満たし難い条件となっている。例えば、字幕の場合は、複雑背景上に

重畳されていることが多く、また映像に特有の色の滲みで文字が繋がってしまうなど、1文字切り出しの精度を低下させる要因が極めて多い。

【0007】

このような印刷文書に限らず、通常、映像中にも、従来の字幕検出法では検出できない文字部分も少なからず存在する。例えば、インクで描かれた紙板をカメラで撮影するフリップ上の文字は、必ずしも特別に高輝度ではなく、そのため境界との輝度差も高いとは云えない。更に、最近の字幕挿入機器の高性能化に伴って、文字の色彩や形の自由度も高くなり、例えば、黒色などの低輝度の文字を縁取りして利用するケースも増えている。また、流れて消えていくタイプの字幕も少なくなく、字幕が必ずしも静止しているとも限らない。

【0008】

また、文字認識に基づくキーワード検索では、文字認識が言語や書体に依存するため、多国語あるいは異書体文字の検索を行なうためには、それぞれの言語や書体に適した認識手法を用意する必要がある。特に、映像中には、ニュースや紀行・語学番組等をはじめ、海外の様々な国の言葉が現れ、それらをキーワードにして検索を行ないたいケースも少なくない。しかし、各言語や書体の種類に合わせて、最適な文字認識手法を逐一導入することは、無駄が多い。

本発明の目的は、従来技術では検出が困難な字幕やフリップ等の文字列を検出可能にすると共に、それによって検出した文字列を、言語や書体に依存することなく共通の枠組みで検索可能にする映像中の文字検索方法及び装置を提供することにある。

【0009】

【課題を解決するための手段】

上記の目的を達成するため、本発明の映像中の文字検索方法は、映像を入力する工程と、入力した映像のフレーム画像から文字領域を形状の特徴に基づいて検出する工程と、該文字領域の第一の画像特徴を抽出する工程と、文字入力手段によって入力された検索対象の文字列を画像として描画する工程と、該文字列画像から第二の画像特徴を抽出する工程と、該第一の画像特徴と該第二の画像特徴とを照合して一致度を求める工程と、一致度が得られた文字列を含む該文字領域を

出力する工程とを有することを特徴とする。

【0010】

このように、一文字一文字ではなく、文字の繋がった列が文字列画像として捉えられ、その文字列画像の形状に基づいて文字領域が検出されるので、従来困難であった高い輝度や大きな輝度差を持たない字幕やフリップの文字領域の検出が可能になる。また、検出がフレーム画像毎に行なわれるので、静止せず流れていく字幕等の文字領域を検出することができる。更に、検出した文字領域に対して文字列画像の有する画像特徴という共通の枠組みで照合が行なわれるので、精度の良い検索が可能になると共に、言語や書体に依存しない検索が可能になる。

【0011】

【発明の実施の形態】

以下、本発明に係る映像中の文字検索方法及び装置を幾つかの図面に示した発明の実施の形態を参照して更に詳細に説明する。

【0012】

図1は、本発明の映像中の文字検索装置の一例を示した概略ブロック図である。

図1において、1は、検索処理を実行するコンピュータ、2は、コンピュータ1の出力画面を表示するCRT (Cathode Ray Tube) 等のディスプレイ装置、3は、コンピュータ1に対する外部からの命令を入力する、キーボードをはじめ、マウスやタッチパネル等のポインティングデバイスによる入力装置である。コンピュータ1へは、映像装置4からの映像信号5がビデオ入力器6を経て入力される。

【0013】

映像装置4は、地上波放送や衛星放送、ケーブルテレビなどの放送番組を受信するためのチューナー装置、もしくは光ディスクやビデオテープ等に記録された映像を再生するための装置である。ビデオ入力器6は、映像装置4からの映像信号5をデジタル画像データ並びにデジタル音声データに逐次変換する。映像装置4から出力される映像が最初からデジタル化されている場合には、アナログからデジタルへの変換は不要になるが、ここでコンピュータが扱いやすいようなフォー

マットに変換を行なうこともできる。

【0014】

コンピュータ1の内部において、ビデオ入力器6からのデジタル画像データは、インタフェース8を介してメモリ9に入り、メモリ9に格納されている検索処理プログラムを実行するCPU (Central Processing Unit) 7によってフレーム画像毎に処理される。また、処理の必要に応じて、各種情報やデータを外部情報記憶装置10に蓄積することができる。メモリ9には、以下に説明する映像中の文字検索処理をコンピュータ1に実行させるプログラムが格納されると共に、その処理によって作成される各種のデータが格納され、必要に応じて参照される。

【0015】

以下では、上記ハードウェア、特にコンピュータ1のCPU7によって実行される、本発明による検索処理のソフトウェアフローの詳細を説明する。

【0016】

本発明では、字幕中の文字を、認識によってコード化された記号としてではなく、字体、すなわち字の持つ独特の形状を利用して検索を行なう。具体的には、次のような手順で操作する。

【0017】

まず、本発明の利用者（ユーザ）は、検索しようとする文字列を、通常のテキスト検索と何ら変わりなく、入力装置3を使ってコンピュータ1に入力する。アルファベットであれば、キーボードのキーの一打一打が文字の入力になり、日本語であれば、さらにローマ字変換やカナ漢字変換等を組み合わせることによって漢字仮名交じり文を入力することができる。

【0018】

キーワードが入力されると、コンピュータ1は、そのキーワードに対応する文字フォントを一字ずつ読み出して描画し、文字列画像を作成する。文字フォントは、ビットマップ形式のデータでも、ベクトル形式のデータでも、どちらでも構わない。検索は、この文字列画像をキーとした類似画検索として行なわれる。即ち、字幕中に画像的に相類似する部分を含むものが検索結果として出力される。

【0019】

現在のコンピュータのオペレーションシステムでは、様々な言語を統一の枠組みで扱えるようになっており、任意の言語の文字列画像を作成することは極めて容易である。そして、本発明の文字検索は、この文字列画像を使って行なうものであり、上述の従来技術による1文字単位での照合ではないので、一文字ずつの切り出しが不要で、切り出し精度の影響を受け難い。また、照合する画像のサイズを比較的大きく取れるので、複雑背景上の文字の場合で、背景の一部が文字として混入されてしまった場合でも、統計的に無視できるようになり、見逃しの少ない安定した照合が可能になる。

【0020】

本発明は、大きく分けて2つの処理に分けることができる。一つは、映像を解析し、字幕やフリップなどの文字領域を検出し、その特徴を抽出する前処理であり、もう一つは、抽出した特徴と、ユーザが入力したキーワードの特徴とを照合し、一致するものを見つける検索処理である。

【0021】

まず、前処理である、字幕検出及び特徴抽出の処理について説明する。字幕の特徴量は、(1)高輝度、(2)エッジ集中、(3)一定期間静止、を利用した手法だけでは検出できない場合が少なからず存在することを既に述べた。そこで、本発明では、もっと文字に普遍な特徴を考える。

【0022】

文字は、歴史的に筆やペン等で書かれ継がれてきた経緯もあり、多くの言語で共通して、線の組み合わせによる構造を持っている。すなわち、文字の存在する部分には、線が密集する傾向がある。しかも、その線の幅は、文字の大きさによって決まる特定の範囲内に収まっていることがほとんどである。本発明では、この特徴を、文字の存在を特定する条件に用いる。また、線の色は単色か、或いは緩やかに変化する模様（グラデーション模様）であることも特徴として利用する。これによって、従来は検出できなかったタイプの字幕が検出可能になる。

【0023】

上記の形状に着目した字幕検出並びに特徴抽出のフローチャートを図2に示す

。ステップ200で各種変数の初期化を行なった後、ビデオ映像から1フレーム画像を入力する（ステップ201）。続くステップ202で、映像に特有の文字の滲みを抑えるためのフィルタ処理を行なう。

【0024】

アナログ信号の映像は、本質的に境界が滲み易いだけでなく、低解像度のテレビ受像機で文字を滑らかに見せるためにアンチエリアシング (anti-aliasing) 等のわざと滲ませる処理を行なっている場合が多い。滲みは、境界を成す2つの画素の間に、両画素の輝度の中間的な輝度を持つ画素が割って入っている状態であり、隣接画素間の輝度差が小さくなり、エッジを検出することが困難になる。これをフィルタで除去する。具体的な処理方法については後述する。

【0025】

次に、ステップ203では、特定の幅を持つ線が抽出され、ステップ204で、それらが特定の領域に集中して存在しているかどうかを検定し、集中している領域を抽出する。そして、抽出された領域が、文字領域として妥当かどうかを改めて検査し（ステップ205）、妥当とみなされれば、領域中に含まれる文字部分即ち文字列の画像特徴を抽出する（ステップ206）。

【0026】

映像中に現れる字幕は、人間が視認できるだけの期間は少なくとも画像中のどこかに存在し続けるので、前回のフレームで抽出した文字と同じかどうかを検査して、新規に現れたものだけを選択的に記憶する（ステップ207）。以上の処理をステップ201に戻って、検索対象とする映像が終わるまで繰り返す。

【0027】

次に、上記で抽出した文字特徴と、ユーザが入力したキーワードの特徴とを照合し、一致するものを見つける検索処理について図3を用いて説明する。

【0028】

まず、変数等の初期化をステップ300で行ない、続くステップ301で、パーソナルコンピュータやワードプロセッサ等で広く用いられている各種の入力方法を用いて、ユーザから任意の文字列を取得する。この文字列を、予め用意した文字フォントを用いて、ビットマップ画像として描画する（ステップ302）。

こうして得られた画像から文字列の特徴を抽出し（ステップ303）、上記のステップ207で記憶してある文字列特徴の全てに対して一致度を求める（ステップ304、305）。一致度が高い順番にランキングして一覧表示することで、検索結果とする（ステップ306）。

【0029】

以下では、上記フローチャートの各処理の実現方法について、さらに詳細に説明する。

【0030】

まず、図2のステップ202のフィルタ処理においては、具体的には、例えば、図4に示す3×3のフィルタを用いる。図は、中央の画素P5が、P1からP9までの画素P5に隣接する9個の画素の輝度値を総合することによって決まることを示しており、9画素中の最大値と最小値を求め、いずれか近い値にP5の輝度値を補正する。これを式で表すと次のようになる。即ち、

$$P_{\max} = \text{MAX} \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9\}$$

$$P_{\min} = \text{MIN} \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9\}$$

$$\begin{aligned} \text{If } P_5 - P_{\min} < P_{\max} - P_5 \quad \text{then } P_5 &\leftarrow P_{\min} \\ \text{else } P_5 &\leftarrow P_{\max} \end{aligned}$$

となる。

【0031】

画像サイズが画素の数で幅w、高さhで表される場合、原点(0, 0)を画像左上の頂点とした場合、x座標が1からw-2まで、y座標が1からh-2までの範囲のすべての座標について、P5の補正を行なう。境界部の滲み除去が目的なので、高速処理のため、境界部、すなわちエッジとなっている部分のみに着目して補正処理を行なうのでも構わない。必要ならば、上記の例の3×3ではなく、5×5や7×7のサイズのフィルタにしてもよい。このフィルタにより、滲みによって境界に発生していた中間的な輝度値は、境界を成すいずれかの画素の輝度値に統一され、境界が明確になる。

【0032】

次に、ステップ203の特定線幅部抽出処理及びステップ204の特定線幅部

集中領域抽出処理について詳細に説明する。特定線幅部抽出処理の概略を図5に示す。特定の幅を持った線だけを選択的に抽出する方法としては様々な方法が考えられるが、厳密に線を抽出して幅を求めるのではなく、もっと簡易的に、特定の幅を満たす線の存在に結びつく特徴がある部分を抽出する。

【0033】

具体的には、垂直・水平各方向について、予め指定する範囲の輝度の差で、予め指定する範囲の長さ分だけ連続して繋がっている等輝度画素列を抽出し、その各画素に印を付ける。図中の○印を付けられた画素が、水平方向の等輝度画素列を構成する画素を意味し、×印は同じく垂直方向の等輝度画素列を示している。垂直・水平方向とも、3画素及び4画素の長さが指定の範囲である。

【0034】

これら○×印の付いた画素の多い領域に文字が存在する。領域の特定には、印刷文書の読み取り認識の分野で一般的に用いられている投影分布法を利用することができる。但し、文字は、縦線と横線が複雑に絡み合って構成されることが多いので、文字のある領域には、水平・垂直各方向の等輝度画素列が同時に集中して存在する。すなわち、○印の画素と×印の画素とが同時に集まっている。この特徴を利用することで、特定性能の向上が期待できる。

【0035】

そこで、フレーム画像において、垂直・水平各方向の画素間隔による各走査毎に○印の存在数をカウントした投影分布と、×印の存在数をカウントした投影分布をそれぞれ作成し、各行ごとに値を掛け合わせた値からなる投影分布を作成し、これをもとに文字領域を検出する。○印と×印のいずれとも存在しなければ、投影分布の値は0になり、文字領域の特定性能が高まる。なお、この乗算の際には、画素単位による走査では、文字領域であっても○×が同時に現れない場合も多いので、余裕をもって複数走査分の値の総和同士を掛け合わせるようにした方がよい場合がある。

【0036】

なお、等輝度画素列を構成する画素は、予め指定する範囲の輝度の差である他、輝度そのものが予め指定する範囲にあるようにすることが可能である。

【0037】

特定線幅部抽出処理のステップ203のフローチャートを図6及び図7に示す。この処理では、処理対象のフレーム画像のサイズに合わせた二次元配列Map(i, j)を初期値0で用意し、水平・垂直の2方向について、特定長の等輝度画素列が検出される毎に、Map(i, j)上の対応する位置の値を0以外の値に変える。

【0038】

まず、ステップ400では、Map(i, j)の初期化を行ない、ステップ401では、y方向のループカウンタをリセットする。続くステップ402では、x方向のループカウンタと、等輝度画素列の始点を初期値にセットする。

【0039】

ステップ403では、x方向に隣接する2画素の輝度の差の絶対値を求め、閾値TH1以下であれば、何もせずステップ407に進む。閾値TH1よりも大きければ、ステップ404に進む。ステップ404では、最後に記憶した、等輝度画素列の始点と現在地点との差が閾値TH2よりも大きく、かつ閾値TH3よりも小さいかどうかを比較し、条件を満たす場合に、この画素列が所定幅の線の一部である可能性が高いとして、この画素列に対応する位置のMap配列に1を代入する（ステップ405）。そして、画素列の始点を現在地点に更新して（ステップ406）、x方向のループカウンタを1進め（ステップ407）、フレーム画像の幅wだけ（ステップ408）、以上を繰り返す（ステップ403～408）。更に、y方向のループカウンタを1進め（ステップ409）、これをフレーム画像の高さh分だけ（ステップ410）、以上を繰り返す（ステップ402～410）。

【0040】

同様にして、図7に示すように、y方向の等輝度画素列も求めることができる（ステップ411～420）。このときの等輝度画素列の分布は、410までで求めたMap配列に2を足し込むことで記憶する（415）。即ち、Map配列の値が0のときは、x, y両方向とも等輝度画素列に含まれないことを意味し、1は、x方向のみ、2はy方向のみ、そして、3は両方向ともに等輝度画素列に含まれることを示すことになる。

【0041】

次に、ステップ 2 0 4 の特定線幅部集中領域の抽出は、基本的には Map 配列に 0 以外の値が入っている画素が集中している領域を見つければよい。Map 配列からの文字領域の抽出には、既に述べたように投影分布を利用する。

【 0 0 4 2 】

図 8 に示すように、画面 6 0 0 中の横書きの字幕を検出する場合は、水平方向に投影分布 6 0 2 をとると、字幕の存在する部分に投影分布のピークが現れやすい。これによって、字幕の存在する y 方向の始点と高さが得られる。そして、このようにして得られた y 軸の範囲の領域について、同様に垂直方向の投影分布をとり、x 方向の始点と幅とを確定することになる。縦書きの字幕の場合は、まず垂直方向の投影分布をとって、x 方向の始点と幅を最初に得、次いで y 方向の始点と高さを確定すればよい。

【 0 0 4 3 】

具体的には、図 9 に示すフローチャートに従って処理を行なう。まずステップ 5 0 0 では、投影分布 ProjY の初期化を行ない、続くステップ 5 0 1 で y 方向のループカウンタのリセットを行なう。ステップ 5 0 2 では、x 方向のループカウンタ及び x, y 方向それぞれの等輝度画素列の数をカウントする変数 x_line, y_line に初期値 0 を入れる。

【 0 0 4 4 】

全体を画素単位走査でスキャンし、x 方向の等輝度画素列であれば、x_line に 1 を加え、y 方向の等輝度画素列であれば、y_line に 1 を加える（ステップ 5 0 3 ～ 5 0 8 ）。

【 0 0 4 5 】

そして、x_line と y_line を掛け合わせたものを投影分布 ProjY に代入する（ステップ 5 0 9 ）。フレーム画像の高さだけ（ステップ 5 1 0, 5 1 1 ）、以上を繰り返す（ステップ 5 0 2 ～ 5 1 1 ）。ステップ 5 0 9 により、x, y 両方向の等輝度画素列が存在しないと、投影分布は十分に大きな値を取らないようになる。

【 0 0 4 6 】

ステップ 2 0 5 では、抽出された文字領域が最終的に文字領域として妥当かどうかの判定を行なう。例えば、検出された領域の高さや幅が極端に小さい場合は

、文字ではない可能性が高い。そのため、予め設定した閾値よりも高さや幅が小さい場合には、文字領域とはみなさず破棄する。

【0047】

また、エッジの多い複雑背景上の文字領域の場合、文字と背景との正確な切り分けは、文字を認識して形を確定しない限り困難である。このような場合でも、可能な限り背景を除去するため、文字領域を余裕を持って大きめに取り囲む矩形の線上の各画素を起点にして、文字領域の内側を探索し、同じ色や輝度の画素を背景として除去する。文字は周囲との間に明確な境界を持つので、背景として一緒に除去されてしまうことはない。

【0048】

もし、文字領域の多くが背景として除去されてしまったなら、それは画像中の物体や背景の模様が文字として誤認識された結果であり、文字領域として妥当性を欠いていたと判定できる。したがって、背景除去処理後に、一定面積以上の文字部分が残っていなければ破棄する。この妥当性判定処理を、何種類かの閾値範囲の輝度、線幅、縦書き・横書きの組み合わせで得られる各文字候補領域について行ない、妥当とされたものを文字領域とする。

【0049】

ステップ206の文字列特徴抽出では、字幕として抽出された文字領域の画像特徴を抽出する。字幕画像そのものをそのまま特徴として記憶し、テンプレートマッチング等の古典的な画像照合を行なうこともできるが、記憶するデータ量が大きくなり、また演算効率の面でも問題が多い。文字の照合にあたっては、字幕とテンプレートとの間でサイズが異なるため、サイズを様々に変えて最も一致度の高いときの値を見極めてから、最終的な照合結果としなければならないが、二次元のフリーサイズの照合は、計算量が非常に多く、大量の照合を必要とする検索用途には不向きである。

【0050】

そこで、本実施例では、図10に示すような、二値化した文字画像について、一定間隔で縦のエッジ数をカウントして横に並べた一次元の特徴列が採用される。x, yの二次元のうち、y方向の情報を伸縮によって値が変化しないエッジ数

で代表させることによって、一次元の特徴列で文字列の画像特徴を表現する。

【0051】

なお、縦書きの文字画像の場合は、一定間隔で横のエッジ数をカウントして縦に並べた一次元の特徴列を採用することができる。

【0052】

以上のようにして得られた映像中の文字列特徴は、図3に示したステップ304において、ユーザが入力した文字列の特徴(ステップ303)との一致度の計算が行なわれる、即ち、特徴照合が行なわれる。照合の際は、1次元のみの伸縮照合で良い。1次元の伸縮照合方法については、例えば、DP (Dynamic Programming) マッチングに代表される様々な高速化手法を適用することができる。

【0053】

この1次元特徴列の一つ一つの値は、単なるエッジの数に過ぎず、個々の情報量は小さいが、その順列組み合わせの数は文字列の長さに比例して膨大になり、検索に意味のある単語であれば十分な文字列長が見込めるので、高い特定性能が期待できる。更に、単純図形の文字ばかりから成る文字列でも正しく検索するために、例えば、単なるエッジ数ではなく、エッジが、縦(|)・横(-)・右斜め(/)、左斜め(\)のうち、いずれの種類の線の境界として形成されたものかを特徴量の一つとして加えてもよい。

【0054】

図11は、上記で述べた文字検索方法を利用する際のユーザインタフェースの一例を示したものである。図11において、700は、検索結果出力画面の例であり、705は、検索キーワードの入力画面の例である。ユーザは、キーワード入力用のテキスト入力領域706に、任意の文字列をキーボード等を利用して打ち込む。正しく文字列を入力できたことを確認後、検索ボタン709を押下することで、検索処理が開始され、キーワード文字列の検索結果が700のように表示される。検索をとりやめるときは、キャンセルボタン710を押下する。

【0055】

画面700には、検索結果として、キーワード文字列と合致する文字列を含む字幕を持つフレーム画像701と一致度704の組み合わせが、一致度の高い順

番に一覧表示される。これによって、検索した字幕が現れているシーンが一目でわかる。一般に、一度に沢山の結果を見たい場合が多いので、表示されるフレーム画像は小さなサムネール表示とする必要がある。しかし、それでは字幕が縮小されて読みにくくなり、検索結果が正しいかどうかの判断も困難になる。そのため、字幕部分だけを別枠702で拡大して表示する。また、検索されたキーワード文字列を枠703で囲んで、強調表示することも可能である。なお、枠703は、特徴抽出によって判定された文字列を囲むので、文字列に対するその位置が背景や字体等によって多少変わる特徴がある。また、フレーム画像と一緒にタイムコードなどのアドレス情報を表示することで、そのシーンの映像中での位置が分かるようにしてもよい。更に、一覧中に表示される情報の密度を向上させるため、上位に存在する文字列検索結果が1画面に複数存在する場合は、1画面のみにまとめてより下位のランキングのものが繰り上げ当選で表示されるようにしてもよい。

【0056】

また、本発明では、その原理上、入力されたキーワードを文字列画像に変換する際に選ばれたフォントと全く異なる書体の場合、字幕から正しく検出できない場合がある。そこで、フォントを入力欄707で指定することができるようにしてもよい。主要なフォントの種類は限られているので、ボタン708を押下することで選択肢を示し、その中から選べるようにしても構わない。文字認識に基づく検索の場合は、認識処理アルゴリズムに修正が必要となるが、本発明では、このように単にフォントの切り換え操作のみで柔軟に対応することができる。

【0057】

図12は、検索キーワードの入力画面705を用いずに検索を行なう場合の一例である。画面800に表示されている字幕をポインティングデバイス等でポイントすることで、その字幕と同じ文字列が現れている字幕をすべて検索する。この方法では、キーワードが最初から画像で与えられるので、文字列画像に変換する必要がなく、フォントも不要になる。例えば、知らない文字が字幕に現れているが、どうやらそれが番組のキーワードになっているらしい、という場合に、その文字が入力できなくても、そのキーワードが現れているシーンをすべて見つけ

ることができる。未習得の言語の語学番組を効率的に見る手助けにもなる。

【0058】

また、本発明によれば、予め特定のキーワードを指定しておけば、そのキーワードが映像中に現れた瞬間に自動的に通知することもできる。例えば、「速報」、「台風」、「地震」或いは「不通」などの言葉を記憶しておけば、緊急の情報を察知することができる。

【0059】

また、同様に、広告映像の場合には、「www」や「http」等の関連情報のアドレスを示す定型的なフレーズを自動検出することもでき、その画面だけを選択的に記録しておいて、後の参照に備えたりすることが可能になる。特にインターネットのドメイン名やURL (Uniform Resource Locator) などのアドレス情報の場合、広告を行なう企業や団体は数が限定されているので、予め企業や団体のアドレス情報のテンプレートパターンを用意しておき、それとマッチングを行なって企業や団体を特定した上で、それに関係づけられた文字コード化されたアドレス情報を用い、実際にインターネットブラウザ等で自動接続して、そのホームページを閲覧することも可能である。この場合、ユーザにとっては、テレビ放送に合わせて、ブラウザの接続先が自動的に変更されているように見える。或いは、ユーザがマウス等でそのアドレス文字領域をクリックした場合に、はじめて接続するようにしてもよい。

【0060】

更に、本発明は、映像データベースの分野において、何ら人手による索引付けなしにテキスト入力によるシーン検索を可能にする手段として利用することができるだけでなく、人手によるキーワード付け作業の効率化にも貢献できる。索引付けを行なう場合に、特定のキーワードを付けるべきシーンを本発明によって見つけ、索引として登録していく。これによって、次回からは、索引データに対する純粋なテキスト検索のみで高速に検索を行なえるようになる。

【0061】

また、本発明の処理対象は、テレビ放送の映像に限定されるものではなく、幅広く自然画像にも適用可能であることは云うまでもない。例えば、街中にある看

板等に書かれた文字列も上述の処理の対象にすることができる。旅行等で観光地を訪れた際に撮影したスナップ写真やビデオクリップの中から、看板の文字をキーにして検索を行なうことができる。特に、名所旧跡によくある、地名が書かれた表札や由來說明と一緒に撮影していれば、記憶に残りやすい地名をキーにして過去に撮影した膨大なライブラリの中から、所望の写真やビデオを探し出すことができる。

【 0 0 6 2 】

そしてまた、前述したアドレス情報の検出による自動ネットワーク接続を、このような自然画像を対象に行なうこともできる。例えば、看板等に書かれたアドレス情報を撮影し、それに対して上述の処理を行なうことで、そのホームページを簡単に呼び出すことができる。或いは、この場合、撮影と本発明の処理、そしてホームページの表示とを、手軽に持ち運べる小型の機器の中で一括して行なえることが望ましい。一つの方法は、そのような機器を作ることであるが、画像処理を迅速に行なえるほど高速な演算性能を小型機器の中に搭載することは、技術的もしくはコスト的に困難な場合がある。そのため、携帯電話やネットワーク接続可能な携帯端末等を利用することで、演算負荷の重い処理をネットワーク上のサーバに行なわせることもできる。

【 0 0 6 3 】

図 1 2 に、その一例を示す。携帯電話機 9 0 0 には、カメラ 9 0 1 が搭載されており、任意の看板等を撮影することができる。図 1 3 は、システムのブロック図の一例であり、カメラ 9 0 1 で撮影された画像データを携帯電話機 9 0 0 が送信し、送信基地局 9 2 0 で中継し、通信網 9 2 1 に接続された画像処理サーバ 9 2 2 で、上記のアドレス情報抽出処理を行ない、処理結果を逆の経路を辿って携帯電話機 9 0 0 に返信する。サーバ 9 2 2 には、データ記録装置 9 2 3 が接続されており、アドレス情報のテンプレートデータ等が格納される。このテンプレートデータを常に最新の情報に保守しておくことで、新種のアドレス情報に対しても、携帯電話機のユーザの側では何ら操作を行なうことなく自動対応が可能になる。

【 0 0 6 4 】

【発明の効果】

本発明によれば、文字単位ではなく文字列単位の照合によって検索を行なうので、ビデオ映像中の字幕やフリップ等の文字について、言語を問わず、文字認識なしに、任意の文字列が検索可能になる。また、複雑背景上に文字がある場合でも、即ち背景に多少のノイズがあっても安定して検索を行なうことができる。

【図面の簡単な説明】

【図 1】

本発明に係る映像中の文字検索装置の実施形態の例を説明するための構成図。

【図 2】

本発明に係る映像中の文字検索方法の実施形態における前処理を説明するためのフローチャート。

【図 3】

本発明に係る映像中の文字検索方法の実施形態における検索処理を説明するためのフローチャート。

【図 4】

滲み除去フィルタを説明するための図。

【図 5】

文字領域抽出方法の概要を説明するためのイメージ図。

【図 6】

特定線幅部特徴抽出処理を説明するためのフローチャート。

【図 7】

特定線幅部特徴抽出処理を説明するための図 6 に続くフローチャート。

【図 8】

文字の存在する領域を求める概要を示すイメージ図。

【図 9】

特定線幅部集中領域の抽出を説明するためのフローチャート。

【図 1 0】

文字列の特徴抽出の概要を示すイメージ図。

【図 1 1】

本発明による検索時の操作インタフェースの一例を説明するための図。

【図 12】

タッチパネルを用いた操作インタフェースの一例を説明するための図。

【図 13】

携帯電話機を用いた本発明の利用方法の一例を説明するための図。

【図 14】

携帯電話機を用いた本発明の利用方法を実現するシステム構成の一例を説明するための図。

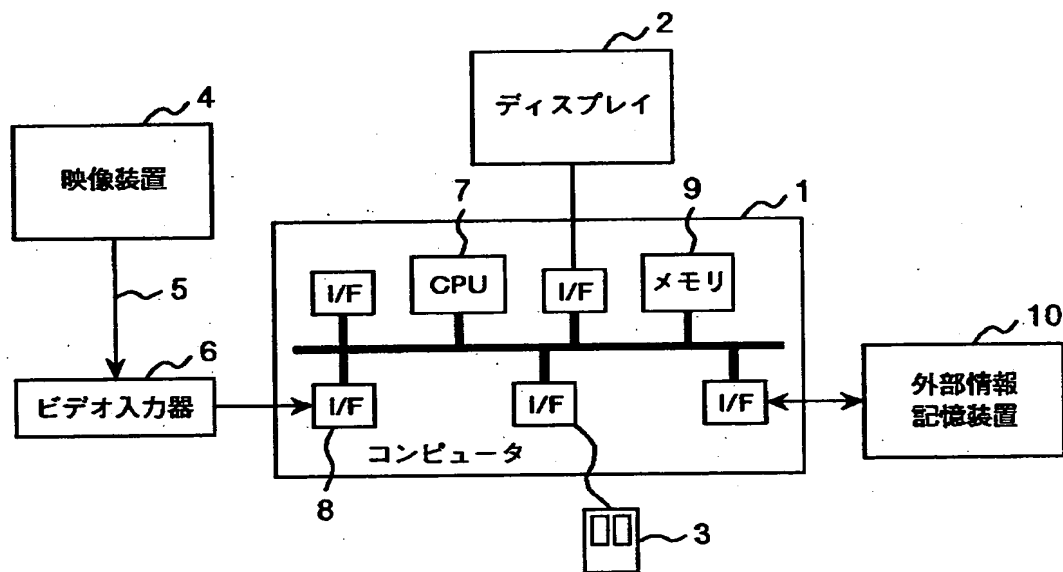
【符号の説明】

1…コンピュータ、2…ディスプレイ、3…入力装置、4…映像装置、5…映像信号、6…ビデオ入力器、7…CPU、8…インタフェース、9…メモリ、10…外部情報記憶装置。

【書類名】図面

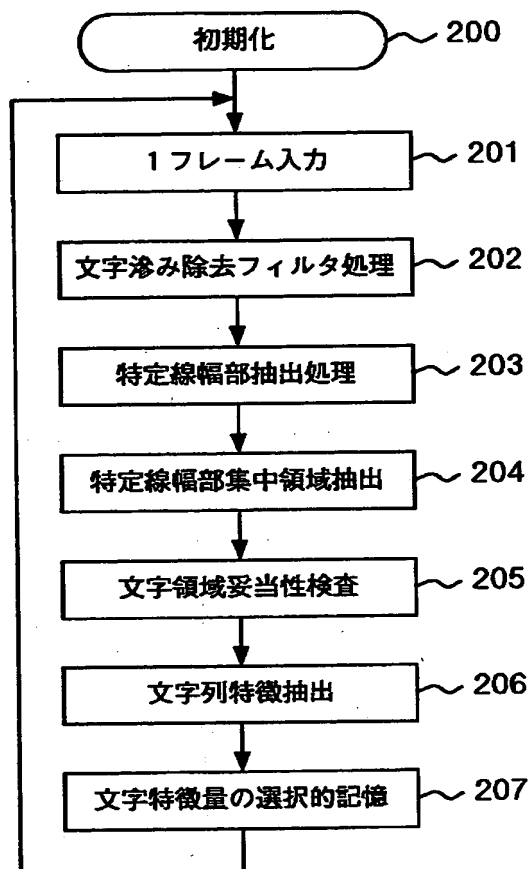
【図1】

図 1



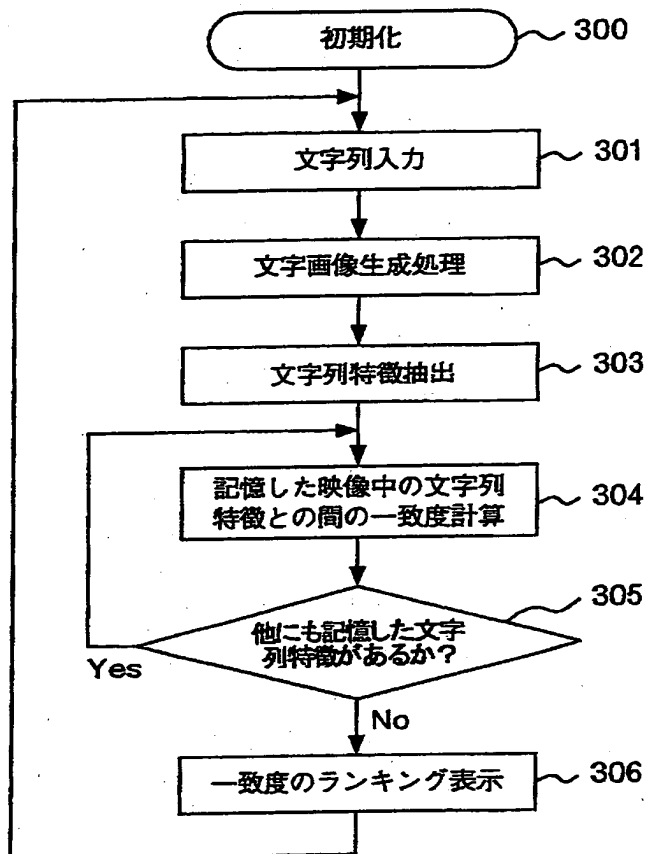
【図 2】

図 2



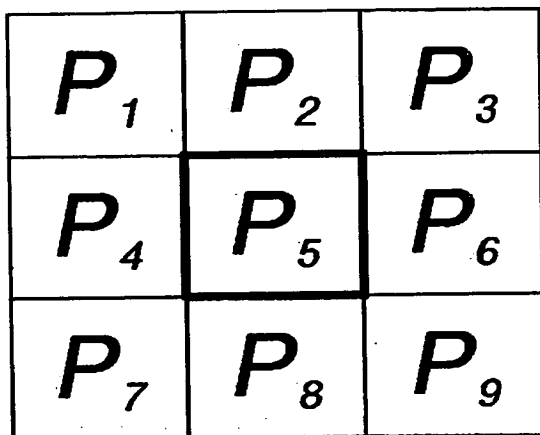
【図3】

図 3



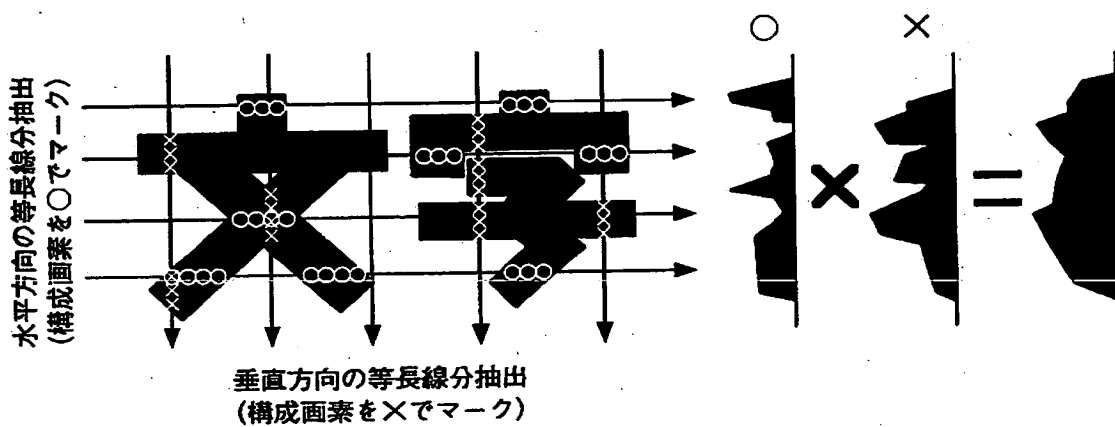
【図4】

図 4



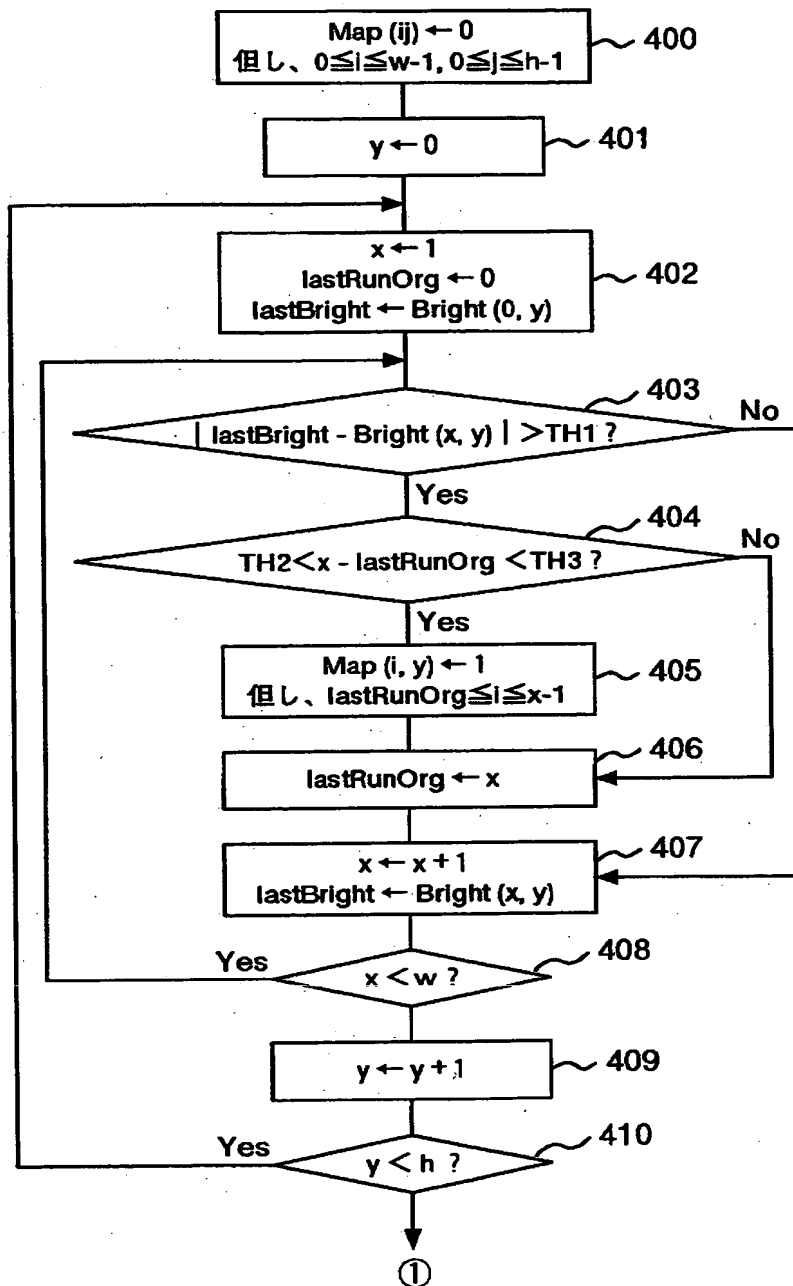
【図5】

図 5



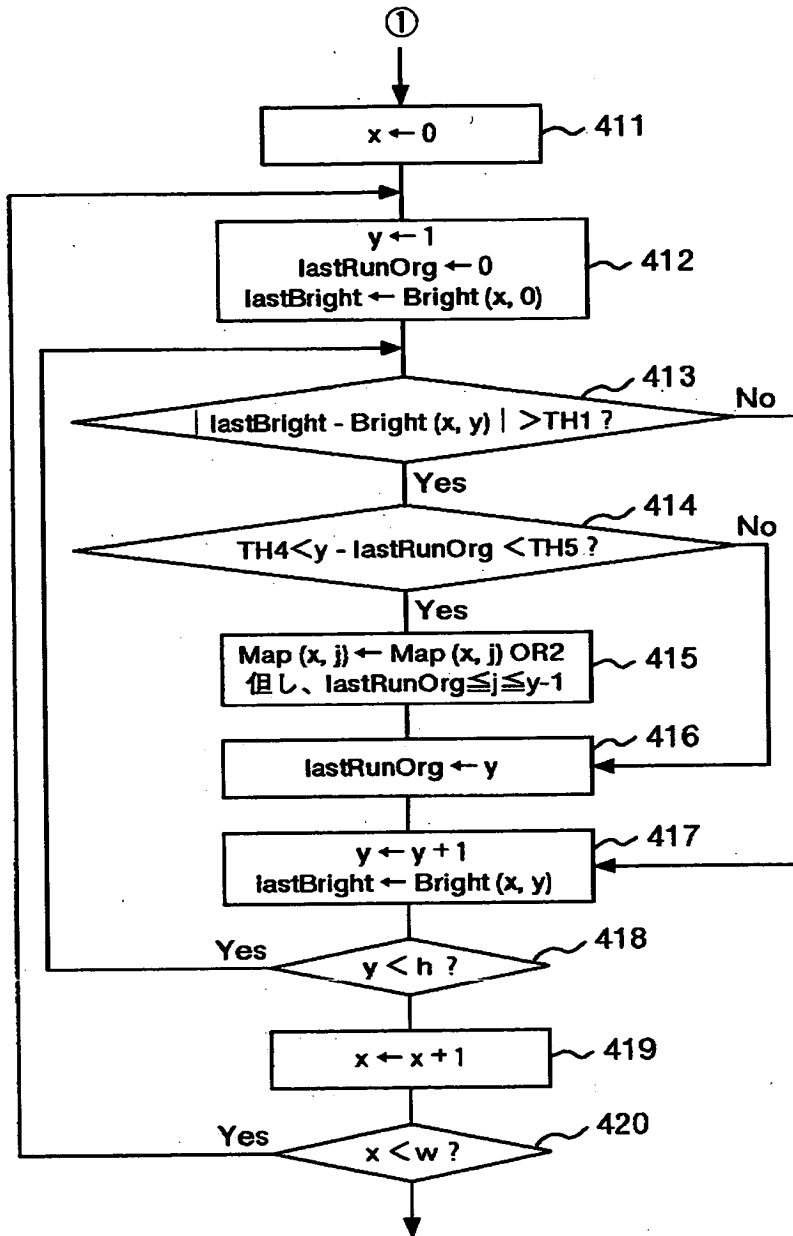
【図 6】

図 6



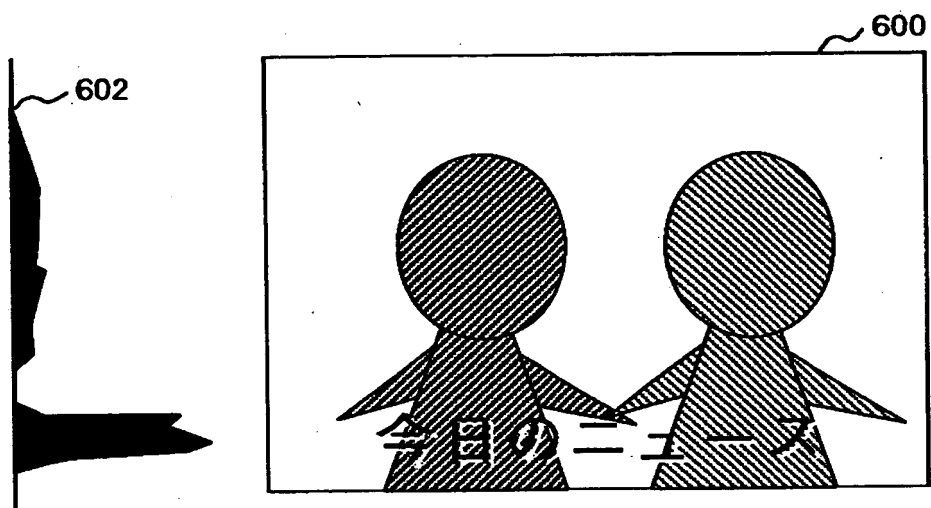
【図 7】

図 7



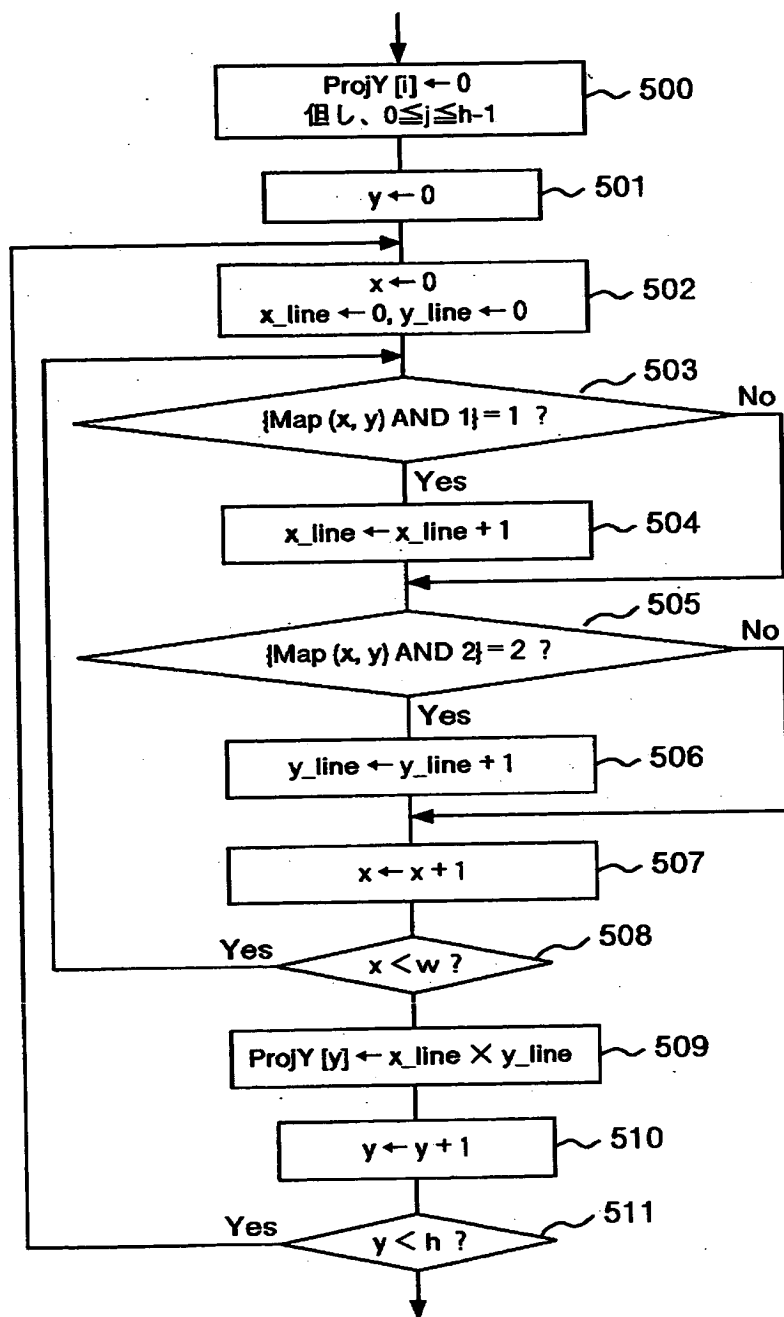
【図8】

図 8



【図 9】

図 9



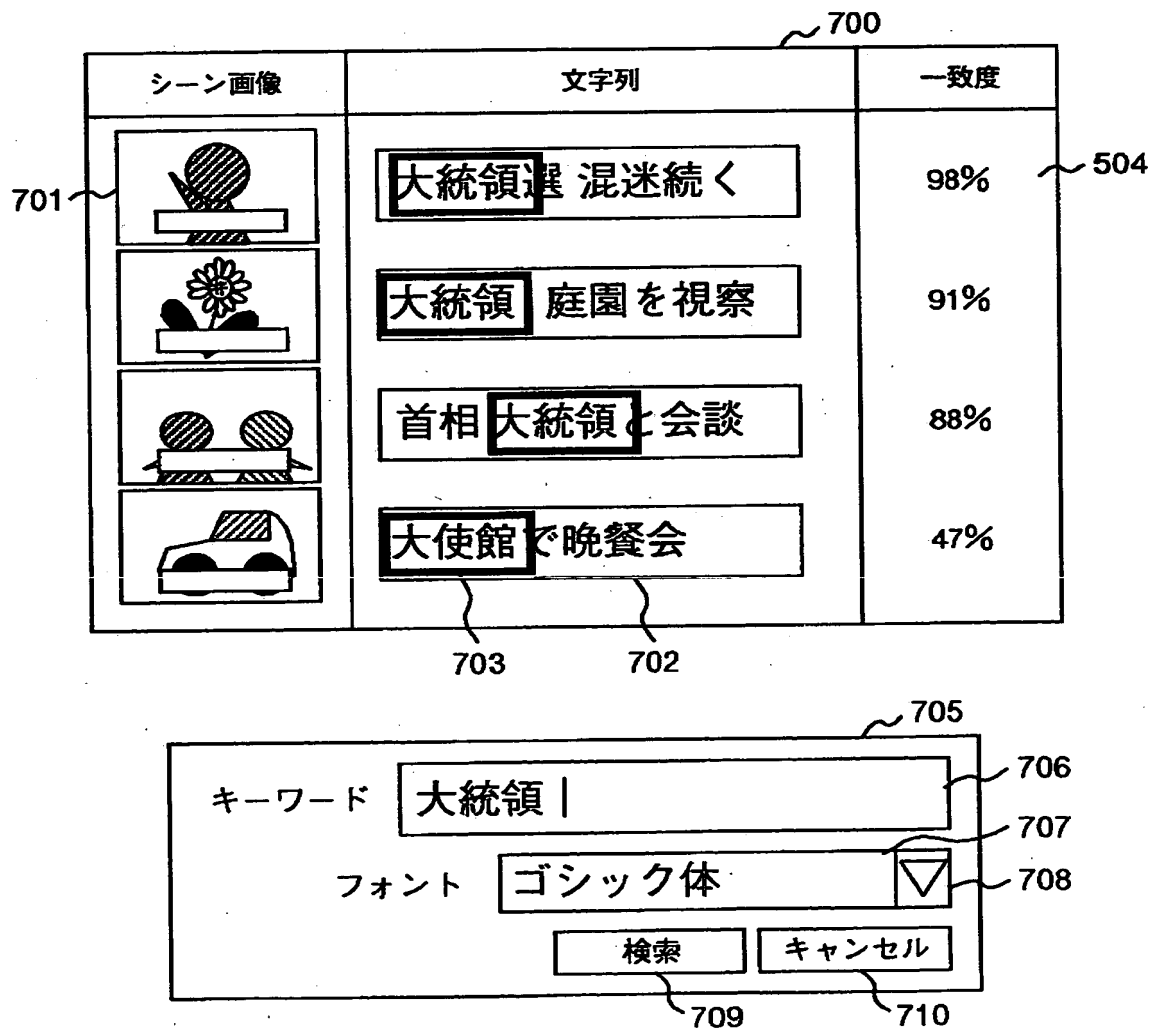
【図10】

図 10

大統領選 混迷続く
 02112011210213104554002212035110213201110

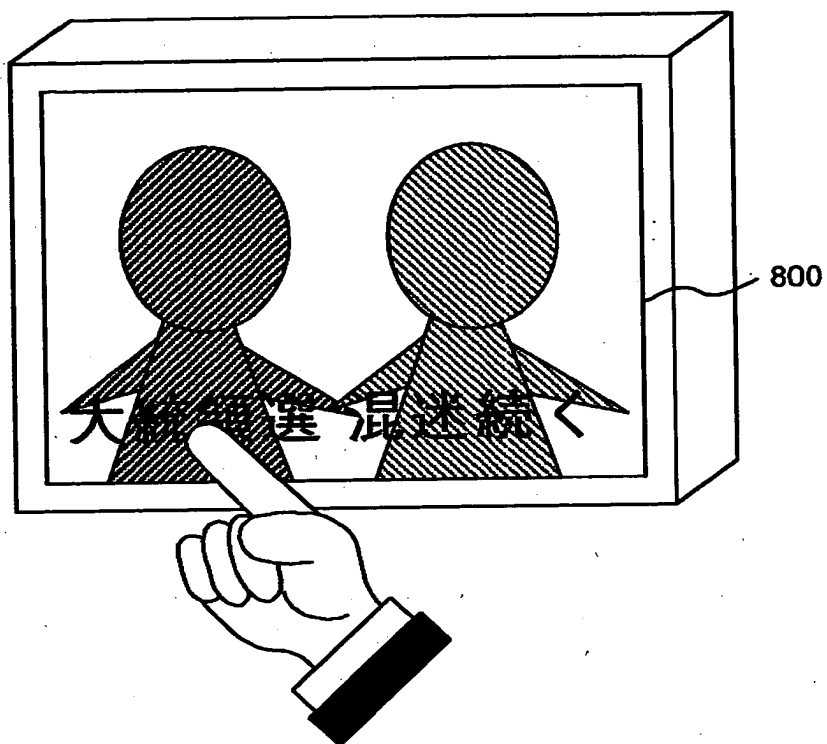
【図11】

図 11



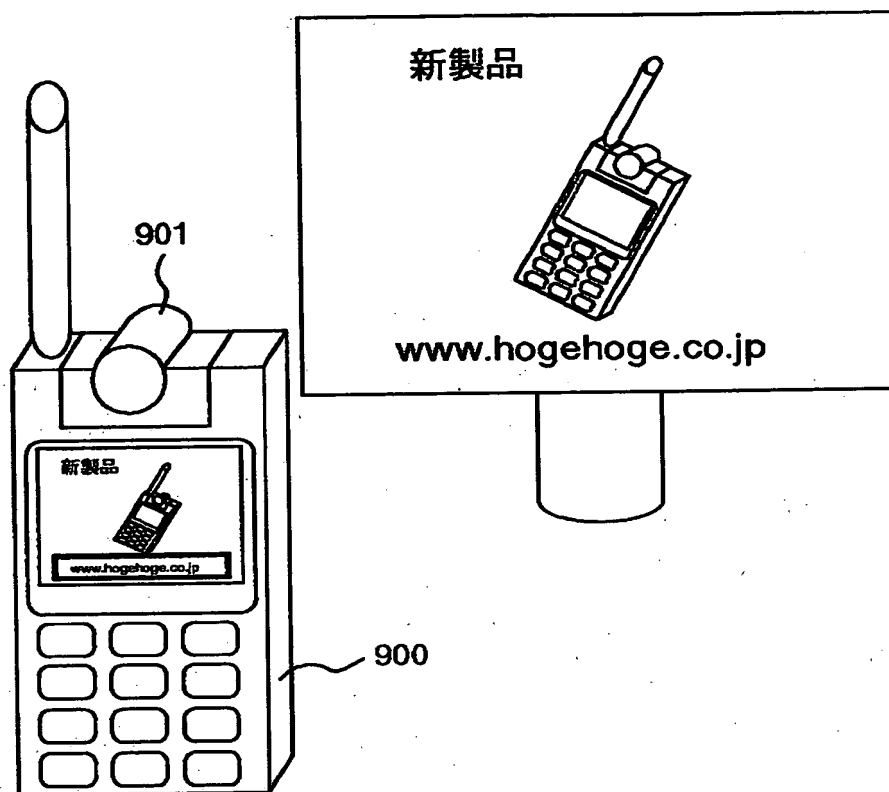
【図12】

図 12



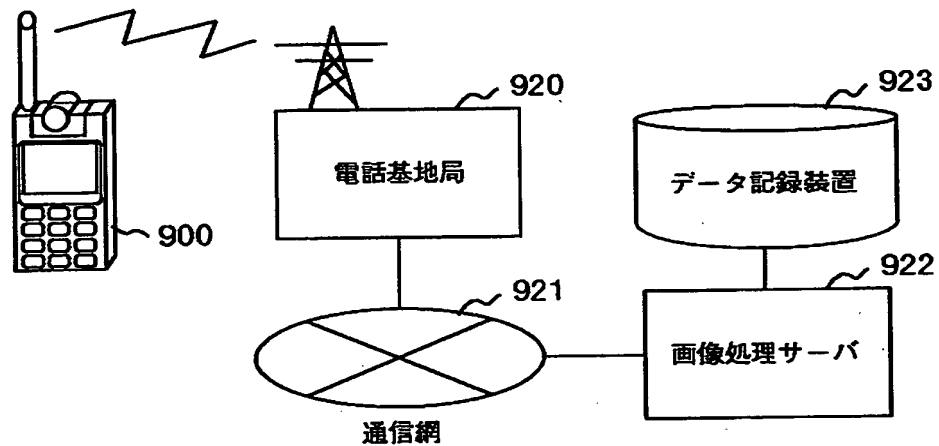
【図 13】

図 13



【図14】

図 14



【書類名】 要約書

【要約】

【課題】従来技術では検出が困難な字幕やフリップ等の文字列を検出可能にすると共に、それによって検出した文字列を、言語や書体に依存することなく共通の枠組みで検索可能にする映像中の文字検索方法及び装置を提供すること。

【解決手段】映像を入力し、入力した映像のフレーム画像から文字領域を形状の特徴に基づいて検出し、文字領域の画像特徴を抽出する。一方、文字入力手段によって入力した検索対象の文字列を画像として描画し、該文字列画像から画像特徴を抽出する。得られた文字領域の画像特徴と文字列画像の画像特徴とを照合して一致度を求め、一致度が得られた文字列を含む該文字領域を出力する。

【選択図】 図3

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台4丁目6番地
氏 名	株式会社日立製作所